Inf. Pauta Fortaleza, CE v. 1 n. 1 jan./jun. 2016

ARTIGO

ONTOLOGIA COMO RECURSO DE PADRONIZAÇÃO TERMINOLÓGICA NO PROCESSO DE RECUPERAÇÃO DE INFORMAÇÃO

ONTOLOGY AS STANDARDIZATION FEATURE TERMINOLOGICAL IN THE PROCESS OF INFORMATION RETRIEVAL

Jorge Janaite Neto
UNESP
Edberto Ferneda
UNESP

RESUMO

O processo de recuperação de informação envolve um acervo documental que deve ser representado por expressões linguísticas e usuários que tentam descrever linguisticamente as suas necessidades de informação a fim de obterem documentos relevantes para satisfazer tais necessidades. Um sistema de recuperação de informação é, portanto, um ambiente linguístico mediador na comunicação entre um estoque de informação e seus requisitantes. Sua eficiência depende de um controle adequado da linguagem de representação dos itens de informação e das requisições dos usuários. Este trabalho apresenta um método de utilização de ontologias na recuperação de informação. Utiliza-se o Modelo Espaço Vetorial como estrutura formal para a representação dos documentos e das buscas dos usuários. Os vetores dos documentos são criados durante o processo de indexação automática no qual uma ontologia fornece novos termos além daqueles extraídos do texto, enriquecendo, assim a representação do documento. O vetor de busca é criado a partir de um processo de expansão de consulta no qual, a partir de inferências em uma ontologia, novos termos são inseridos na expressão de busca inicialmente formulada pelo usuário. Pretende-se com isso uniformizar as representações dos documentos e das buscas, melhorando assim a precisão do processo de recuperação de informação.

Palavras-chave: Recuperação de informação. Indexação automática. Expansão de consulta. Ontologia. Modelos de recuperação de informação.

ABSTRACT

Information Retrieval process involves a collection of documents that must be represented by linguistic expressions and users trying to describe linguistically their information need in order to obtain relevant documents that meet such needs. Therefore, an information retrieval system is a linguistic environment mediating the communication between a stock of information and its users. Its effectiveness depends on adequate control of language for representation of information items and requests of its users. This paper presents a method of using ontologies in information retrieval process. It uses the Vector Space Model as a formal structure for the representation of documents and queries. The documents vectors are created during the automatic indexing process, in which the ontologies provide new terms in order to enrich those representations. The search vector is created from a query expansion process in which, from inferences in ontology, new terms are entered in the search expression initially formulated by the user. The aim is standardize the representations of documents and searches, thus improving the information retrieval process.

Keywords: Information retrieval. Automatic indexing. Query expansion. Ontology. Information retrieval models.

1 INTRODUÇÃO

A recuperação de informação pode ser vista como um processo de comunicação que envolve, por um lado, o conhecimento registrado e representado em um suporte material, os documentos, e, por outro, temos pessoas que buscam por documentos relevantes para satisfazer as mais variadas necessidades de informação. Nesse sentido, Meadow *et al*. (2007, p. 3) afirmam que:

> Information Retrieval is a communication process. In one sense it is a means by which authors or creators of records communicate with readers, but indirectly and with a possibly long time lag between creation of a message or text and its delivery to the IRS user. Sometimes the IRS or librarian conducting a search may pass on information about the probable relevance or value of what is retrieved, thereby adding information to the set of retrieved items. The records of a database are created and assembled without knowledge of exactly who will read them, or under what circumstances. The languages and channels of such a communication system are quite different from other well-known models, such as broadcasting or point-to-point communication.[1]

Pode-se, assim, inferir que um sistema de recuperação de informação é um ambiente linguístico cuja eficiência depende de um controle adequado da linguagem de representação dos itens de informação e das requisições de seus usuários. Insere-se como um agente mediador na comunicação entre um estoque de informação e os seus potenciais requisitantes.

Como mediador de um processo comunicativo, uma das tarefas de um sistema de recuperação de informação é definir uma linguagem (código) comum entre emissor e receptor, entre os documentos e as requisições dos usuários. Na Ciência da Informação, os vocabulários controlados são tradicionalmente considerados uma ferramenta para compatibilizar a terminologia utilizada na representação dos documentos, e o vocabulário empregado pelo usuário na representação de sua necessidade de informação. Cintra (2002) afirma que a construção desses vocabulários visa às atividades de indexação, armazenamento e recuperação de informação.

Para Fujita (2004), os vocabulários controlados (linguagens documentárias) visam representar conceitos significativos de assuntos dos documentos utilizados na fase de indexação e busca. Ainda segundo a autora, proporcionam uma convergência entre a

---

[1] Recuperação de informação é um processo de comunicação. Em certo sentido é um meio pelo qual autores e criadores de registros se comunicam com os leitores, mas indiretamente e possivelmente com um longo intervalo de tempo entre a criação de uma mensagem ou texto e a sua entrega para o usuário de um sistema de recuperação de informação. Às vezes, o sistema de recuperação de informação ou bibliotecário conduzindo uma pesquisa pode passar informações sobre a provável relevância ou valor do que é recuperado, aumentando a informação para o conjunto de itens recuperados. Os registros de uma base de dados são criados e montados sem conhecimento exatamente de quem irá lê-los, ou sob quais circunstâncias. As linguagens e os canais de tal sistema de comunicação são bastante diferentes de outros modelos bem conhecidos, tais como a radiodifusão ou a comunicação ponto-a-ponto. (Tradução nossa).

linguagem do indexador e a linguagem do usuário de um sistema de informação, "[...] já que vários autores podem utilizar diferentes palavras para expressar uma mesma ideia, assim como os usuários podem apresentar diversidade de vocabulário quando da expressão de uma estratégia de busca".

Este trabalho tem por objetivo apresentar um método de utilização de ontologias como um elemento normalizador e unificador da linguagem de representação dos documentos e das buscas dos usuários a fim de melhorar a precisão dos resultados alcançados no processo de recuperação de informação.

2 RECUPERAÇÃO DE INFORMAÇÃO

Em 1951, Calvin Mooers criou o termo *Information Retrieval* e definiu os problemas a serem tratados por essa nova disciplina:

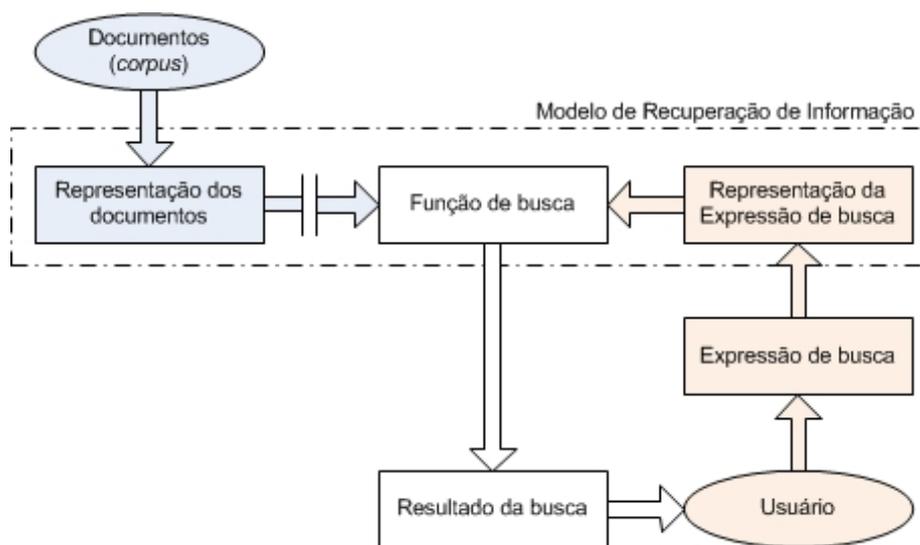
> Information retrieval is the name for the process or method whereby a prospective user of information is able to convert his need for information into an actual list of citations to documents in storage containing information useful to him. It is the finding or discovery process with respect to stored information. It is another, more general, name for the production of a demand bibliography. Information retrieval embraces the intellectual aspects of the description of information and its specification for search, and also whatever systems, techniques, or machines that are employed to carry out the operation. Information retrieval is crucial to documentation and organization of knowledge.[2] (MOOERS, 1951, p. 20).

Para Saracevic (1999), a Recuperação de Informação pode ser considerada a vertente tecnológica da Ciência da Informação, e é resultado da relação desta com a Ciência da Computação.

Recuperar uma informação consiste em identificar em um acervo documental quais documentos satisfazem total ou parcialmente a uma determinada necessidade de informação do usuário. A Figura 1 apresenta uma representação do processo de recuperação de informação:

---

[2] Recuperação de informação é o nome para o processo ou método pelo qual um usuário em potencial de informação é capaz de converter a necessidade de informação em uma lista real de citações de documentos no armazenamento contendo informações úteis para ele. É o processo de encontro ou descoberta no que diz respeito às informações armazenadas. É outro, mais geral, o nome para a produção de uma bibliografia sob demanda. Recuperação de informação abrange os aspectos intelectuais da descrição das informações e sua especificação para a pesquisa, e também quaisquer sistemas, técnicas, ou máquinas que são utilizadas para realizar a operação. Recuperação de informação é crucial para documentação e organização do conhecimento. (Tradução nossa).

Figura 1 – Representação do processo de recuperação de informação.

Fonte: (FERNEDA, 2012, p. 14).

Iniciando a descrição da Figura 1 pelos "documentos", Buckland (1991) define o conceito de "informação como coisa" e argumenta que os documentos de um sistema de informação seriam registros relacionados a coisas ou objetos. Nesses sistemas, informação está vinculada ao objeto que a contém.

A eficiência de um sistema de recuperação de informação depende da forma como esses documentos estão representados. A "representação dos documentos" de um sistema de recuperação de informação tem por objetivo identificar e descrever resumidamente o conteúdo informacional de cada um, permitindo a localização e recuperação.

Em um sistema de recuperação de informação o "usuário" expressa sua necessidade de informação por meio de uma "expressão de busca", composta geralmente por um conjunto de termos. É necessário que a expressão de busca seja representada de forma similar à forma como os documentos foram representados, para que seja possível uma comparação entre essas duas representações.

No centro do processo de recuperação de informação está a "função de busca", que compara as representações dos documentos com a representação da expressão de busca e recupera os itens que supostamente fornecerão informações relevantes. De forma geral, a função de busca calcula o grau de similaridade entre a expressão de busca e cada um dos documentos do *corpus*.

Segundo Ferneda (2012, p. 20), um modelo de recuperação de informação é a especificação formal de três elementos básicos: a representação dos documentos, a representação da expressão de busca e a função de busca. Baeza-Yates e Ribeiro-Neto (2011, p. 58) definem modelo de recuperação de informação como uma quádrupla:

$$[D, Q, F, R(q_i, d_j)]$$

> D é um conjunto composto por visões lógicas (representações) dos documentos no *corpus*;
> Q é um conjunto composto de visões lógicas das necessidades de informação dos usuários;
> F é um *framework* para a modelagem de representações dos documentos, consultas e seus relacionamentos;
> $R(q_i, d_j)$ é uma função de ordenamento *(ranking)* que atribui um número real à relação entre uma representação da consulta q_i de Q e a representação de um documento d_j de D.

Os primeiros modelos de recuperação de informação, os chamados modelos "clássicos",[3] datam das décadas de 1960 e 1970, e as suas principais ideias ainda estão presentes na maioria dos sistemas de recuperação atuais e nos mecanismos de busca da Web. O método apresentado neste trabalho está embasado no Modelo Vetorial, que será apresentado a seguir.

3 MODELO VETORIAL

No Modelo Espaço Vetorial (ou simplesmente Modelo Vetorial), um documento é representado por um vetor numérico onde cada elemento representa o peso, ou relevância, do respectivo termo de indexação na representação do documento. Uma expressão de busca é também representada por um vetor numérico onde cada elemento representa a

[3] Os modelos clássicos de recuperação de informação são: Booleano, Vetorial e Probabilístico. Para a proposta deste trabalho, será detalhado somente o Modelo Vetorial.

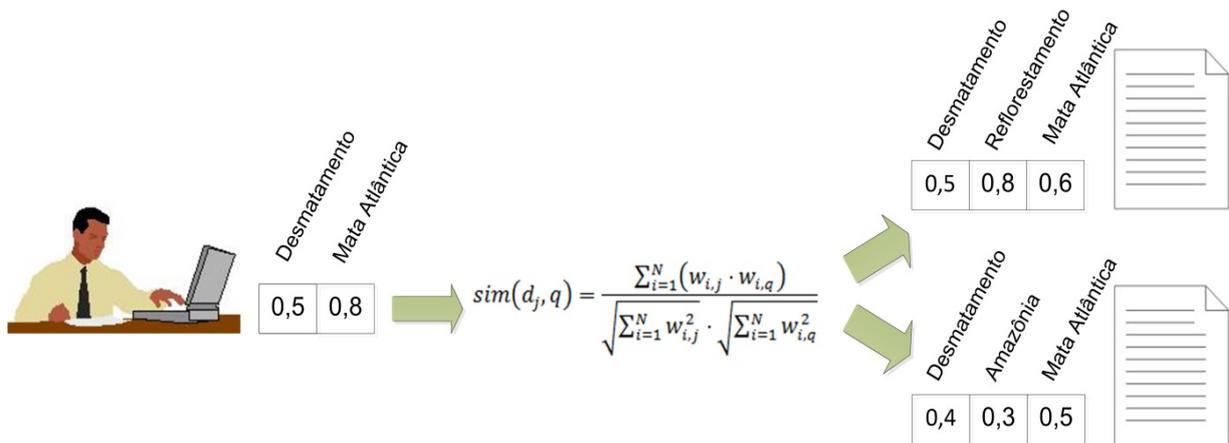
importância (peso) de cada termo na representação da necessidade de informação do usuário (SALTON; WONG; YANG, 1975).

A utilização de uma mesma representação, tanto para os documentos como para as expressões de busca, permite calcular o grau de similaridade entre o vetor que representa uma determinada expressão de busca e cada um dos vetores que representa os documentos. Em um espaço vetorial contendo ***N*** dimensões, a similaridade ***(sim)*** entre um documento $\boldsymbol{d_j}$ e uma consulta $\boldsymbol{q}$ é obtida por meio da seguinte fórmula (SALTON; McGILL, 1983, p. 121):

$$(d_j, q) = \frac{\sum_{i=1}^{N}(w_{i,j} \cdot w_{i,q})}{\sqrt{\sum_{i=1}^{N} w_{i,j}^2} \cdot \sqrt{\sum_{i=1}^{N} w_{i,q}^2}}$$

Onde $w_{i,j}$ é o peso do *i*-ésimo termo do documento d_j e $w_{i,q}$ é o peso do *i*-ésimo termo da expressão de busca q.

A Figura 2 apresenta uma ilustração da representação vetorial de uma expressão de busca e dois documentos.

Figura 2 – Ilustração do Modelo Vetorial.

Fonte: Elaborado pelos autores.

O cálculo da similaridade *(sim)* é efetuado entre o vetor que representa a expressão de busca do usuário e cada um dos vetores representativos de cada documento do *corpus.* Os valores da similaridade *(sim)* entre a expressão de busca e cada um dos documentos são utilizados no ordenamento do resultado de busca.

4 ONTOLOGIA

Uma ontologia pode ser considerada como um vocabulário de representação dos conceitos relacionados a algum domínio do conhecimento (CHANDRASEKARAN; JOSEPHSON; BENJAMIN, 1999). Segundo Jacob (2003, p. 19):

> Ontologias são categorias de coisas que existem ou podem existir em um determinado domínio particular, produzindo um catálogo onde existem as relações entre os tipos e até os subtipos do domínio, provendo um entendimento comum e compartilhado do conhecimento de um domínio que pode ser comunicado entre pessoas e programas de aplicação.

Conforme Jasper e Uschold (1999), uma ontologia pode possuir uma variedade de formas, mas necessariamente incluirá um vocabulário de termos, e alguma especificação de seus significados. Isto inclui definições e uma indicação de como conceitos estão inter-relacionados, o que impõe uma estrutura no domínio e restringe as possíveis interpretações dos termos.

Uma ontologia define os conceitos usados em uma determinada área de conhecimento, padronizando seus significados. Pode ser usada por pessoas, bases de dados e aplicações que precisam compartilhar informações e conceitos de um domínio (DACONTA; OBRST; SMITH, 2003).

A construção de uma ontologia pode ser pensada como uma união de elementos que formam uma estrutura complexa. Classes e subclasses definem um "esqueleto" na forma de uma hierarquia, complementada por propriedades descritivas, propriedades relacionais, regras e axiomas.

Toda classe é caracterizada por seus atributos ou propriedades. Uma subclasse herda as características (atributos) da classe-pai. Uma instância é a materialização de uma classe e representa um conceito ou uma entidade do mundo real. Quando uma classe é instanciada, cada um dos seus atributos pode, então, receber valores que irão individualizar aquele conceito ou entidade. É possível estabelecer regras que impõem restrições e limites às classes e atributos e que se refletem nas suas instâncias.

Portanto, uma ontologia é uma estrutura conceitual que visa representar formalmente os conceitos e suas relações, regras e restrições lógicas de um determinado domínio, e pode ser definida por meio de linguagens processáveis por computadores.

5 INDEXAÇÃO AUTOMÁTICA BASEADA EM ONTOLOGIA

Os primeiros trabalhos sobre indexação automática consideravam o texto de um documento como um elemento autônomo, cuja semântica se resolveria no interior do próprio texto. Em abordagens posteriores começaram a surgir pesquisas que utilizavam algum elemento externo aos documentos para dar suporte à indexação automática. Esses elementos podem ter diferentes níveis de complexidade, podendo variar de simples listas de palavras até tesauros e ontologias.

As ontologias podem ser utilizadas na indexação automática, pois oferecem uma estrutura conceitual e terminológica restrita a um determinado domínio e originalmente representada em linguagens processáveis por computador. A partir de um conjunto de documentos textuais, são identificados e selecionados termos que possam ser mapeados para os conceitos de uma ontologia. Isso permite padronizar o vocabulário e restringir o campo semântico dos termos ao domínio da ontologia, solucionando, assim, possíveis ambiguidades.

O método de indexação automática proposto neste trabalho inicia-se com a extração de um conjunto de termos que represente o conteúdo informacional de cada documento. Para cada termo é atribuído um valor numérico (peso) que expressa a relevância do respectivo termo na representação do documento. A extração de termos e o cálculo de seus pesos são realizados por meio de um método matemático, tal como o método de indexação definido por Salton, Wong e Yang (1975). Esse processo de obter termos que indicam os assuntos tratados por um documento textual se estabeleceu como um campo de pesquisa na Ciência da Computação, denominado "Extração de Informação" *(Information Extraction)* (SARAWAGI, 2008).

Extração de informação é, portanto, a tarefa de extrair informação de forma automática a partir de documentos legíveis por computador. Essa extração pode ser realizada por meio de métodos puramente matemáticos (estatísticos) ou pela utilização de métodos e técnicas de Processamento de Linguagem Natural (GRISHMAN, 1997).

De forma genérica, ao final desse processo poderia se obter, por exemplo, os seguintes termos de indexação e seus respectivos pesos (Figura 3):

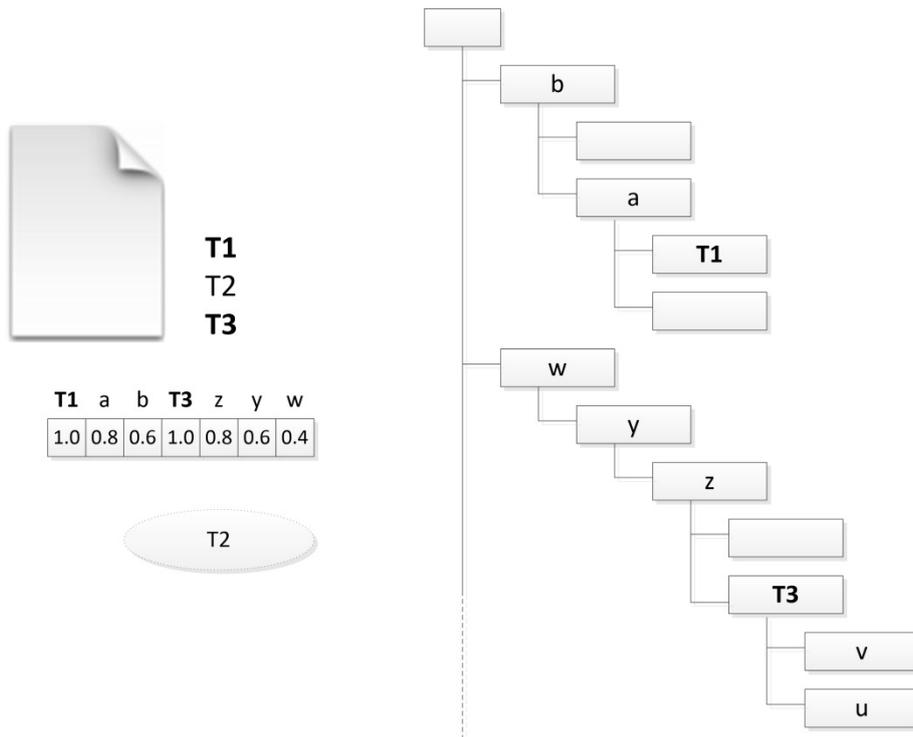
Figura 3 – Termos de indexação e pesos.

Fonte: Elaborado pelos autores.

Nesse exemplo foram extraídos do documento seis termos com seus respectivos pesos. Considerando um parâmetro que define o peso mínimo para que o termo possa ser considerado um termo de indexação, serão utilizados apenas os termos cujo peso seja maior ou igual a 0.8. No caso do exemplo, o documento será representado apenas pelos três temos de maior peso, desconsiderando os termos com pesos menores que 0.8.

Os termos extraídos do documento serão considerados sucessivamente como conceitos centrais da ontologia associada ao *corpus.* A ontologia terá duas funções: (1) expandir o conjunto de termos de indexação de cada documento; e (2) atribuir pesos a cada um dos termos.

No exemplo da Figura 4 é apresentado o documento do exemplo acima e uma ontologia representada genericamente por sua estrutura hierárquica. Verifica-se que apenas os termos **T1** e **T3** possuem relação com conceitos da ontologia. Os demais termos que irão compor o vetor do documento serão derivados desses dois conceitos por meio suas relações.

Figura 4 – Representação vetorial de um documento utilizando ontologia.

Fonte: Elaborado pelos autores.

No exemplo da Figura 4, o termo **T1** será representado no vetor do documento com peso igual a 1. Na ontologia, os termos hierarquicamente mais genéricos a **T1** receberão pesos decrescentes, dependendo da "distância" de **T1.** O termo **a** recebe o peso 0.8, e o termo **b,** 0.6. Considerando agora o conceito **T3,** ele receberá peso 1, pois faz parte da ontologia. Os conceitos **z**, **y** e **w** receberão, respectivamente, os pesos 0.8, 0.6 e 0.4.

O termo **T2** foi descartado por não estar representado na ontologia. Porém, há de se considerar que esse termo foi extraído do texto do documento com um peso de valor significativo. Nesse caso, o termo **T2** será armazenado em um tipo de repositório, formando um conjunto de potenciais conceitos a serem inseridos na ontologia. Se um determinado termo for repetidamente extraído dos documentos, ele poderá ser convertido em um conceito da ontologia.

6 EXPANSÃO DE CONSULTA BASEADA EM ONTOLOGIA

Em um sistema de recuperação de informação, a especificação da expressão de busca é dependente do usuário, que geralmente expressa sua necessidade de informação por meio de um número reduzido de termos ou palavras, não permitindo uma interpretação exata de sua necessidade. Essa dificuldade fez surgirem pesquisas em expansão de consulta *(query expansion).* Expansão de consulta é o termo utilizado para referenciar os métodos e processos que visam melhorar a eficiência da recuperação de informação, baseados no pressuposto de que as consultas definidas pelos usuários muitas vezes não refletem suas reais necessidades de informação. O objetivo principal é adicionar novos termos à consulta inicialmente formulada pelo usuário, a fim de melhorar os resultados obtidos. O conceito de expansão de consulta está relacionado ao conceito mais genérico de reformulação de consulta, que pode envolver também a exclusão de termos de uma consulta inicial.

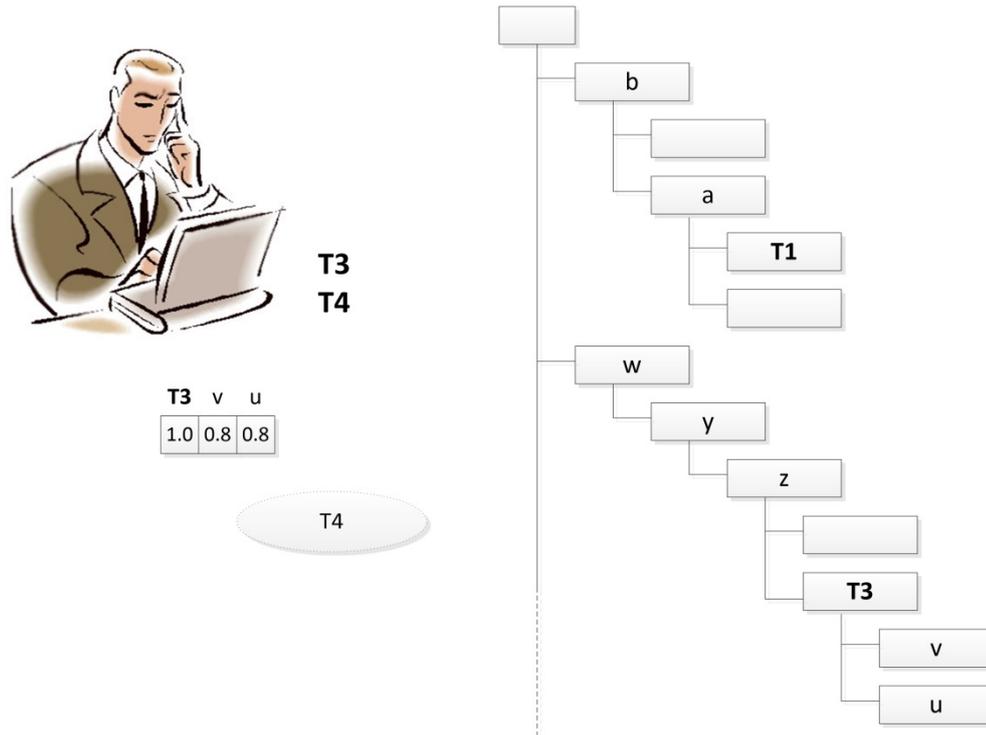
Propõe-se neste trabalho que uma ontologia pode ser utilizada na expansão das consultas inicialmente formuladas pelos usuários, por meio da inserção de novos termos derivados dos relacionamentos entre os conceitos da ontologia.

A partir de uma interface adequada, as ontologias podem servir também como ferramentas para a seleção dos termos que irão compor a consulta inicial do usuário. Isso permite que uma pessoa "leiga" em um determinado domínio ou assunto consiga realizar consultas pertinentes em um sistema de recuperação de informação, ao mesmo tempo em que se familiariza com a terminologia do domínio de interesse.

Como dito anteriormente, no Modelo Vetorial uma expressão de busca é representada por um único vetor numérico no qual cada elemento corresponde à importância do respectivo termo para a descrição da necessidade de informação do usuário.

Antes da execução da busca, o usuário deve selecionar a ontologia do domínio ao qual se refere a sua necessidade de informação. Os termos definidos pelo usuário em sua expressão de busca (consulta) serão utilizados como conceitos centrais da ontologia associada a essa consulta. A ontologia terá duas funções: (1) expandir o conjunto de termos da consulta, acrescentando novos termos provenientes da ontologia; e (2) atribuir pesos a cada um dos termos da consulta.

No exemplo da Figura 5, após o usuário selecionar a ontologia referente ao tema de sua busca, ele utilizou dois termos de busca: **T3** e **T4**. Fazendo-se uma busca na ontologia selecionada, verifica-se que apenas o primeiro termo está representado na ontologia. Assim, no vetor que representará esta consulta, apenas o termo **T3** estará presente com peso igual a 1. O termo **T4** será descartado.

Figura 5 – Representação vetorial de uma expressão de busca utilizando ontologia.

Fonte: Elaborado pelos autores.

Tomando-se **T3** como conceito central da ontologia e considerando os conceitos específicos, derivam-se os termos **v** e **u**, que farão parte da expressão de busca expandida. Ambos os termos receberão o valor 0.8, como exemplificado na Figura 4.

O termo **T4**, que não está presente na ontologia, será armazenado em um tipo de repositório que, dependendo da frequência com que esse termo for utilizado nas buscas dos usuários, poderá ser convertido em um conceito da ontologia.

7 CONSIDERAÇÕES FINAIS

Diante da grande quantidade de informação disponível na Web, recursos de busca e recuperação de informação estão presentes em diversos sites para as mais variadas finalidades. Cotidianamente, utilizam-se tais recursos de forma natural e intuitiva para diversas tarefas rotineiras. Pesquisa-se o menor preço de um determinado produto antes de adquiri-lo em uma loja virtual, que, por sua vez, possui um sistema que auxilia seus clientes na tarefa de encontrar o produto desejado. Nas livrarias online, é possível encontrar obras do autor favorito ou o *best-seller* do momento. Em sites corporativos é muito comum um campo de busca com o qual se podem encontrar informações sobre um assunto de interesse no contexto daquela empresa.

Essas ferramentas ou sistemas apresentam resultados relativamente satisfatórios, pois foram criados para atender a um domínio bastante restrito, no qual os itens de informação são conhecidos, e as buscas podem ser facilmente previsíveis. Nesses “ambientes controlados”, os problemas linguísticos são minimizados, pois utilizam uma terminologia cujo campo semântico está restrito a uma determinada área, um assunto, ou mesmo a um ramo de atividade.

As ferramentas ou mecanismos de busca de propósito geral, tais como o Google e o Bing, têm pretensões universalistas de abarcar toda a informação livremente disponível na Web. A ausência de uma delimitação explícita do contexto semântico dos termos com os quais os documentos e as necessidades do usuário são representados, afeta na precisão dos resultados de busca.

Em um sistema de recuperação de informação existem dois principais elementos de natureza linguística: a representação dos documentos e a representação da expressão de busca. A eficiência do sistema é dependente da correta interpretação dos documentos e das necessidades de informação do usuário, a fim de gerar suas respectivas representações. Além dos aspectos semânticos envolvidos nesse processo, tais representações devem estar formalmente estruturadas para que possam ser utilizadas por um sistema computacional.

No método apresentado neste trabalho, os elementos linguísticos que formam uma ontologia são considerados termos de um vocabulário de domínio, utilizado como ferramenta de padronização terminológica das representações dos documentos e das buscas em um sistema de recuperação de informação. Tais representações utilizam como

base formal o Modelo Espaço Vetorial, que fornece uma base matemática consistente e consolidada.

Uma vantagem do método é a delimitação explícita do contexto no qual o processo de recuperação de informação é realizado. Todo documento faz parte de um *corpus* documental, cujo domínio é definido pela ontologia a ele associada. Os documentos são indexados utilizando o vocabulário de domínio definido pelos conceitos dessa ontologia. Por sua vez, o usuário define o seu domínio de interesse por meio da seleção de uma ontologia, que será utilizada para agregar novos termos à expressão de busca inicialmente formulada por ele. O Modelo Vetorial fornece a estrutura formal de representação tanto para os documentos como para as buscas, o que permite fornecer como resultado uma lista de documentos ordenados pelo grau de similaridade/relevância.

É importante observar no método aqui proposto a similaridade entre o processo de construção dos vetores representativos dos documentos e das buscas. Essa uniformidade permitirá uma economia significativa no código dos programas utilizados em sua futura implementação.

REFERÊNCIAS

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern Information Retrieval.** 2nd ed. New York: Addison-Wesley, 2011.

BUCKLAND, M. K. Information as thing. **Journal of the American Society of Information Science,** v. 42, n. 5, p. 351-360, 1991.

CHANDRASEKARAN, B.; JOSEPHSON, J. R.; BENJAMINS, V. R. What are ontologies, and why do we need them? **IEEE Intelligent Systems,** v. 14, n. 1, 1999.

CINTRA, A. M. M. (Org.). **Para entender as linguagens documentárias.** 2. ed. São Paulo: Polis, 2002.

DACONTA, M. C.; OBRST, L. J.; SMITH, K. T. **The semantic Web:** a guide to the Future of XML, Web Services, and Knowledge Management. Indianápolis: Wiley Publishing, 2003.

FERNEDA, E. **Introdução aos modelos computacionais de recuperação de informação.** Rio de Janeiro: Ciência Moderna, 2012.

FUJITA M. S. L. A leitura documentária na perspectiva de suas variáveis: leitor-texto-contexto. **DataGramaZero:** Revista de Ciência da Informação, Rio de Janeiro, v. 5, n. 4, ago. 2004.

GRISHMAN, Ralph. Information extraction; techniques and challenges. *In*: INTERNATIONAL SUMMER SCHOOL SCIE, 1997, New York. **Proceedings...** New York: Springer-Verlag, 1997.

JACOB, E. K. Ontologies and the Semantic Web. **Bulletin of the American Society for Information Science and Technology,** Apr./May 2003.

JASPER, R.; USCHOLD, M. A. Framework for understanding and classifying ontology applications. In: **KRR5-99.** Stockholm. 1999.

MEADOW, C. T. *et al*. **Text Information Retrieval System.** 3rd ed. London, UK: Elsevier, 2007.

MOOERS, C. Zatocoding applied to mechanical organization of knowledge. **American Documentation,** v. 2, n. 1, p. 20-32, 1951.

SALTON, G.; McGILL, J. M. **Introduction to Modern Information Retrieval.** New York, McGraw-Hill, 1983.

______; WONG, A.; YANG, C. S. A Vector Space Model for Automatic Indexing. **Communications of the ACM,** v. 18, n. 11, 1975.

SARACEVIC, T. Information Science. **Journal of the American Society for Information Science,** v. 50, n. 12, p. 1051-1063, 1999.

SARAWAGI, S. Information Extraction. **Foundations and Trends in Databases,** v. 1, n. 3, 2008.

TÁLAMO, M. F. G. M.; LARA, M. L. G.; KOBASHI, N. Y. Contribuição da terminologia para a elaboração de tesauros. **Ciência da Informação,** v. 21, n. 3, 1992.

SOBRE OS AUTORES

Jorge Janaite Neto
Analista de Informática da UNESP/Marília. Graduado em História pela UNESP/Marília. Especialista em Planejamento, Implementação e Gestão de Educação à Distância pela UFF e em Redes de Computadores pela UTFPR.
E-mail: janaite@gmail.com

Edberto Ferneda
Professor do Departamento de Ciência da Informação da UNESP/Marília. Bolsista Produtividade em Pesquisa CNPq - Nível 2. Doutor em Ciências da Comunicação pela USP. Pós-doutor pela UFPB.
E-mail: ferneda@marilia.unesp.br

Recebido em: 08/05/2016; **Revisado em:** 05/06/2016; **Aceito em:** 06/06/2016.

Como citar este artigo

JANAITE NETO, Jorge; FERNEDA, Edberto. Ontologia como recurso de padronização terminológica no processo de recuperação de informação. **Informação em Pauta**, Fortaleza, v. 1, n. 1, p. 30-45, jan./jun. 2016.